Acta Psychologica 170 (2016) 112–126

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Involvement of Spearman's *g* in conceptualisation versus execution of complex tasks

Ellen L. Carroll [1], Peter Bright *

Department of Psychology, Faculty of Science & Technology, Anglia Ruskin University, East Road, Cambridge CB1 1PT, UK

ARTICLE INFO

Article history:
Received 1 September 2015
Received in revised form 24 June 2016
Accepted 27 June 2016
Available online xxxx

Keywords:
Spearman's *g*
Fluid intelligence
Response inhibition
Task complexity
Task modelling
Chunking

ABSTRACT

Strong correlations between measures of fluid intelligence (or Spearman's *g*) and working memory are widely reported in the literature, but there is considerable controversy concerning the nature of underlying mechanisms driving this relationship. In the four experiments presented here we consider the role of response conflict and task complexity in the context of real-time task execution demands (Experiments 1–3) and also address recent evidence that *g* confers an advantage at the level of task conceptualisation rather than (or in addition to) task execution (Experiment 4). We observed increased sensitivity of measured fluid intelligence to task performance in the presence (vs. the absence) of response conflict, and this relationship remained when task complexity was reduced. Performance-*g* correlations were also observed in the absence of response conflict, but only in the context of high task complexity. Further, we present evidence that differences in conceptualisation or 'modelling' of task instructions prior to execution had an important mediating effect on observed correlations, but only when the task encompassed a strong element of response inhibition. Our results suggest that individual differences in ability reflect, in large part, variability in the efficiency with which the relational complexity of task constraints are held in mind. It follows that fluid intelligence may support successful task execution through the construction of effective action plans via optimal allocation of limited resources.

1. Introduction

Strong correlations between performance on tests of working memory capacity (WMC) and fluid intelligence (*g*) are well established (e.g., Ackerman, Beier, & Boyle, 2002; Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Süß, Oberauer, Wittmann, Wilhelm, & Schulze, 2002; Unsworth, Redick, Lakey, & Young, 2010). The mediating factors in this relationship, however, are not fully understood. Traditionally, the working memory (WM) system has been presented as a mental workspace associated with the concurrent storage and processing of information; Baddeley and Hitch's (1974) multicomponent WM model, for example, comprises domain-specific storage buffers and a central executive. Complex span tests, which typically assess memory for words or digits in the face of a demanding interleaved task are among the best measures of WMC and are also sensitive to variations in fluid intelligence. In contrast, simple span tests, which do not encompass additional processing demands, are typically weakly correlated with measures of WMC and intelligence (e.g., Conway, Kane, & Engle, 2003; Daneman & Carpenter, 1980; Engle, Tuholski, Laughlin, & Conway, 1999; Turner & Engle, 1989). This finding has led some authors to argue for the central role of processing (e.g., executive attention; Conway et al., 2003; Engle et al., 1999) in driving the correlation between WMC and *g*. Subsequent evidence, however, supports the central role of short-term storage (e.g., immediate recall of memory for numbers, letters, or visual arrays; Chuderski, Taraday, Nęcka, & Smoleń, 2012; Colom, Abad, Quiroga, Shih & Flores-Mendoza, 2008; Colom, Flores-Mendoza, Quiroga, & Privado, 2005).

The executive attention account of inter-individual differences in WMC (e.g., Engle & Kane, 2004; Kane, Conway, Hambrick, & Engle, 2007) claims that individuals with low WMC have relatively limited capacity for controlling goal-directed attention, and for resolving response conflict, compared to individuals with high WMC. According to this view, high WMC individuals typically produce fewer errors on tasks such as the classic Stroop (1935) test (e.g., Kane & Engle, 2003; Long & Prat, 2002) because they possess relatively greater capacity for directing attention to naming the colour, and for resolving the competition between eliciting the prepotent (but incorrect) response of reading the word and producing the appropriate response of naming the colour in which the word is written. On the basis of this executive attention account the memory maintenance and retrieval theory of WMC has been developed (e.g., Unsworth & Engle, 2007a, 2007b; Unsworth & Spillers, 2010) which claims that high WMC individuals are better at both maintaining relevant information in primary (working) memory,

* Corresponding author.
E-mail address: peter.bright@anglia.ac.uk (P. Bright).

[1] Present address: Division of Anaesthesia, School of Clinical Medicine, University of Cambridge, Box 93, Addenbrooke's Hospital, Hills Road, Cambridge CB2 0QQ, UK.

http://dx.doi.org/10.1016/j.actpsy.2016.06.011
0001-6918/

and at using appropriate retrieval cues to retrieve information from secondary (long-term) memory when required. Other researchers stress that response inhibition is the single most important factor in individual differences in WMC (Hasher & Zacks, 1988; Lustig, May & Hasher, 2001; May, Hasher, & Kane, 1999). The claim is that individuals with high WMC are better at restricting WM access to task-relevant information, resolving response conflict, and inhibiting dominant but inappropriate responses. High WMC individuals therefore perform better on the Stroop because they are better able to limit WM access to the relevant task component (naming the colour) and at inhibiting the dominant but inappropriate tendency to name the word. In the context of the maintenance and retrieval view, this theory would also explain how conflict is resolved between the inappropriate stimulus-response mapping of reading the word held in secondary memory and the relevant but less prepotent mapping of stating the colour held in primary memory. The three accounts are not mutually exclusive and share overlapping theoretical claims, but they also incorporate distinct and testable predictions (Redick, Calvo, Gay, & Engle, 2011) as outlined below.

Performance on the Stroop is usually considered to reflect capacity for response inhibition and significant correlations between performance on the Stroop and psychometric intelligence have been reported (e.g., Dempster, Corkill & Jacobi, 1995; Polderman et al., 2009; Salthouse, Atkinson, & Berish, 2003). Functional magnetic resonance imaging (fMRI) studies indicate that the anterior cingulate is recruited in conditions of response conflict (Kerns et al., 2004), and by performance on tasks with high *g*-loadings such as Raven's Advanced Progressive Matrices (Gray, Chabris & Braver, 2003). Cognitive models of “general ability” (e.g., Das, 2002) and prefrontal cortex (Roberts & Pennington, 1996) also highlight the importance of inhibition in intelligence. Nevertheless, studies based on factor analysis have produced inconsistent findings. For example, Salthouse et al. (2003) report a strong relationship ($r = 0.73$) between their composite measures of inhibition and fluid intelligence in a large sample of adults ($N = 261$). Conversely, on the basis of evidence suggesting that executive functions—specifically, inhibiting prepotent responses, shifting between tasks/mental sets, and updating the contents of WM—are correlated but separable (e.g., Miyake et al., 2000; Miyake & Friedman, 2012), Friedman et al. (2006) observed that their composite measures of inhibition ($r = -0.11$) and shifting ($r = -0.08$) did not load significantly onto their fluid intelligence construct whereas WM updating did ($r = 0.74$; WM updating was operationalized by tasks that required the adding and deleting of information held in WM: keep-track, letter-memory, spatial 2-back). Other studies have demonstrated that WMC is not related to the ability to resist interference or dual-task coordination (e.g., Oberauer, Lange & Engle, 2004). These findings indicate that the correlation between response inhibition and intelligence is not straightforward, and therefore that interaction with some other task component(s) may be of critical importance in driving the relationship.

Redick et al. (2011) compared the executive attention, maintenance and retrieval, and inhibition theories of WMC in the context of go/no-go task performance. The authors compared two go/no-go tasks – a *simple* task involving a “go” response to one letter and a “no-go” response to all other letters (with a reverse mapping in another block), and a *conditional* task involving a go response to two letters conditional on the current target being different to the last. Differences between individuals with high/low WMC were observed only in the conditional task, such that high WMC individuals performed better on both target trials (target letters meeting the conditions for a go response) and lure trails (target letters meeting the conditions for a no-go response). Further, performance was significantly correlated with WMC in the conditional task only. These findings were interpreted in the context of the maintenance and retrieval account of WMC, with only the conditional task requiring active monitoring and updating of stimulus-response mappings, and the retrieval of the appropriate goal-relevant response. Redick et al. (2011) argue that if inhibition or executive attention were fundamental aspects of WMC, differences between individuals with high and low spans would also be observed in the simple task, because a prepotent response must be inhibited or response conflict resolved in both tasks.

An alternative view is that if attention is given to maintaining and updating the stimulus-response mappings, reduced attention is available for resolving the conflict associated with the no-go requirement, producing more error on these trials. Consistent with the notion of shared but limited resource availability for processing and storage requirements, research has shown that *anti*-saccade (Mitchell, Macrae, & Gilchrist, 2002; Roberts, Hager, & Heron, 1994) and motor response inhibition (Hester & Garavan, 2005) capacities decline with increasing WM load. Studies directly addressing storage versus processing accounts of the driving force in inter-individual differences in the WMC–*g* relationship emphasise the overarching importance of storage. For example, Colom et al. (2008) claim that simple short-term storage (i.e., memory for numbers/letters/visual arrays) accounts for a large proportion of the relationship between WMC and *g*, and that although attention control, WM updating, and mental speed are independently correlated with *g*, these relationships disappear when short-term storage is controlled for. In a related study, Chuderski et al. (2012) found that their storage latent factor (comprising memory for visual arrays, monitoring of relations among stimuli, and updating information in WM) accounted for 70% of the variance in measured fluid intelligence. For their three processing measures, only attention control, and neither resistance to interference nor response inhibition, was correlated with fluid intelligence (accounting for 25% of the variance in fluid intelligence), yet, when storage was controlled for, this correlation disappeared.

A visual change detection study (Fukuda, Vogel, Mayr, & Awh, 2010) has claimed that the number of representations that can be held in WM is highly correlated with fluid intelligence but that the resolution with which stimulus representations are stored in WM is largely independent. Nevertheless, observations by Duncan, Emslie, Williams, Johnson, & Freer (1996) indicate that the relationship between fluid intelligence and WM cannot be explained on the basis of a straightforward storage function. In their letter monitoring task, participants were able to recall all task requirements after task completion, but the sensitivity to fluid intelligence was explained by differences in the capability for responding appropriately to those requirements. Failure to produce the appropriate response (referred to as “goal neglect”) was largely restricted to participants scoring > 1 SD below the sample mean on the Culture Fair test of fluid intelligence (Cattell, 1971; Cattell & Cattell, 1973).

Duncan et al. (2008) presented evidence that the efficiency with which a task is cognitively modelled or held in mind may be of more fundamental importance to the involvement of Spearman's *g* than the real time processing demands associated with the task. Across a series of computer based experiments, incorporating a variation of the task presented here (Bright, 1998), the authors showed that the form in which instructions were presented was the primary factor predicting both the level of goal neglect and the size of correlation between goal neglect and Spearman's *g*. Thus, although increased task complexity did not increase the level of neglect of task demands, an additional “dummy” requirement which had no impact on what participants were required to do during actual task execution increased level of neglect and the strength of the performance–*g* correlation. This pattern of results has been replicated in children using a slightly simplified version of this feature match task (Roberts & Anderson, 2014). On the basis of their findings Duncan et al. (2008) claim that the ability to attend to a complex “task model – a working memory description of the relevant facts, rules, and requirements used to control current behaviour” (p. 140) is fundamental to individual differences in *g* (see also Bhandari & Duncan, 2014; Dumontheil, Thompson & Duncan, 2010; Duncan, Schramm, Thompson, Dumontheil, 2012).

In the present study we explore the relationship between participants' effective modelling of task demands and Spearman's *g* in the context of other candidate “risk factors” for the recruitment of *g*. In

Experiment 1 we investigate the relationship between response inhibition and *g* by comparing performance–*g* correlations (as well as performance scores) between a task that requires the inhibition of a prepotent response (termed *RI Present*) and a task that requires an alternative, rather than an inhibited, response (termed *RI Absent*) in older participants. In Experiment 2 we determine whether performance on the RI Present task replicates in a younger group of participants. In Experiment 3 we examine the importance of task complexity (vs. response inhibition) in performance and performance–*g* correlations by manipulating the number of task demands in RI Present and RI Absent conditions. In Experiment 4 we explore the importance of task modelling in *g* by directly comparing performance–*g* correlations on the original RI Present task across blocks of trials in which frames requiring response inhibition are either present or absent. In this final experiment we also explore the sensitivity, to task performance and *g*, of the form in which task rules are presented to participants; specifically, the same task information is presented either as a set of two rules or a set of four rules in order to clarify the relationship between WM load, task performance and *g* within the context of a task in which the real time execution demands are held constant.

2. Experiment 1: response inhibition versus no response inhibition

In Experiment 1 we examined risk factors for the recruitment of *g* in task performance. We kept the complexity of task instructions constant and focused on real-time processing demands. Pairs of coloured shapes matching on zero (neither colour nor shape), one (colour or shape), or two (colour and shape) dimensions were rapidly presented on a computer screen (see Fig. 1). One group of participants—the RI Present group—were required to make one of two key press responses (a “go” response) to single-matching pairs on the basis of whether the left or the right shape contained a tick, but to suppress a response (“no-go”) to double-matching pairs (and to also ignore neutral non-matching pairs). Single-matching pairs were presented with higher frequency (30%) than double-matching pairs (7.5%). Therefore, a go response was the dominant response to items matching in colour and items matching in shape; items matching in both colour *and* shape encompass two active response tendencies which must be inhibited for correct performance. Another group of participants—the RI Absent group—completed an almost identical task which differed only in the type of response made to double-matching pairs. The response inhibition demand was removed for this group; instead, participants were required to press the relevant key (left or right) and produce a verbal response. The complexity of this task—participants need to decide if items match, on how many dimensions, and on which side a response should be made—is arguably greater than the stop-signal (Friedman et al., 2006; Salthouse et al., 2003) and go/no-go (Redick et al., 2011, simple version) tasks that have previously been reported to show no link between fluid intelligence and behavioural or motoric type inhibition. We predicted, albeit tentatively, that the task that contained response inhibition requirements, and the specific task items requiring response inhibition (critical errors, see Table 1), would be more sensitive to variations in fluid intelligence.

2.1. Method

2.1.1. Participants

Sixty-nine neurologically healthy adults, aged between 57 and 72 years, were recruited from the MRC Cognition and Brain Sciences Unit panel of volunteers. The RI Present group comprised 38 adults (15 males, 23 females) aged between 57 and 71 years ($M = 63.8$, $SD = 4.6$). The RI Absent group comprised 31 adults (10 males, 21 females) aged between 58 and 72 years ($M = 65.1$, $SD = 4.8$). Culture Fair error (max = 46) was equivalent across the RI Present ($M = 17.71$, $SD = 6.35$) and RI Absent ($M = 16.74$, $SD = 6.31$) groups, $t(67) = 0.63$, $p = 0.529$, $d = 0.153$. All participants gave informed consent and were paid to participate. Our primary reason for sampling from an older group of adults was due to well-established evidence suggesting that cognitive abilities decline (e.g., Hasher, 1997; Rabbitt, 1993;

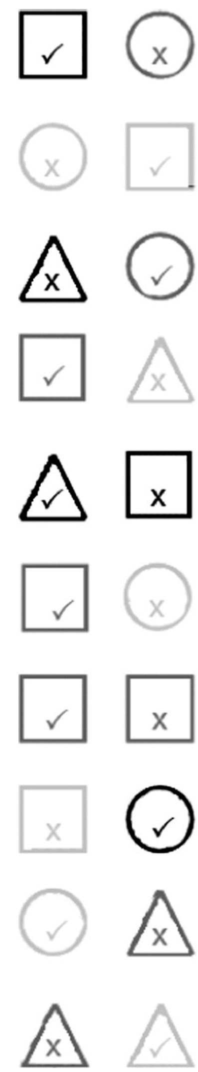

Fig. 1. Representation of a typical RI Present trial in the colour-shape match task. Black, mid grey, and light grey objects are presented in blue, red, and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Definitions of performance measures.

Performance measure	Definition
Critical error (RI Present)	Response to double-matching (NO-GO) pair
Critical error (RI Absent)	Missed ‘double’ (key press and verbal) response to double-matching pair
Miss	No response to single-matching (GO) pair
Hand error	Response to incorrect side of single-matching pair
False positivek	Response (key press) to non-matching (NEUTRAL) pair
False positivev (RI Absent)	Response (verbal) to single- or non-matching pair
Criterion fail	Miss or hand error to first two single-matching pairs
OR	Miss to final single-matching pair
OR	Three or more false positives
OR	Critical error
Reaction time	Time taken (ms) to respond to single-matching pair

Note. All motor responses were made with the index and middle fingers of the right hand.

Salthouse, 1992) and become more variable (e.g., Morse, 1993; Rabbitt, 1993) in later years. Accordingly, older adult performance was assumed to be more sensitive to the demands imposed by the task.

2.1.2. Tasks and procedure

2.1.2.1. Test of 'g': culture fair. Participants first completed Cattell's Test of 'g': Culture Fair, Scale 2, Form A (Cattell, 1971; Cattell & Cattell, 1973; hereafter termed the Culture Fair). The Culture Fair loads highly onto g at $r = 0.81$ and comprises four subtests that measure novel problem-solving ability using geometrical figures in a set amount of time: series completions, classifications, matrices, and topological relations.

2.1.2.2. Colour-shape match task. After completing the Culture Fair, participants performed the colour-shape match task, programmed using PsyScope software (Cohen, MacWhinney, Flatt, & Provost, 1993). This task has previously been employed, in slightly modified form, by Duncan et al. (2008). Pairs of coloured outline shapes, one containing a tick in the centre and the other containing a cross, were presented in the centre of a high resolution colour monitor with a white screen background. Objects varied along two dimensions: colour (red, blue, green) and shape (circle, square, triangle). Across pairs, all combinations of colour and shape were possible. Each object was 12.7 mm × 12.7 mm and the distance between objects in a pair was 7.4 mm edge to edge; participants sat comfortably and viewed these stimuli without precise control of viewing distance (approximately 0.5 m). Pairs were a mixture of non-matching (objects of different colour and shape), single-matching (objects of same colour *or* shape), and double-matching (objects of same colour *and* shape). In each trial, READY was presented on screen, followed by a blank-screen interval of 1500 ms, followed by 10 pairs which were presented for 1200 ms each with a 200 ms blank-screen interval between each pair (see Fig. 1). Of the 10 pairs, three were single-matching, which were always two of the first five frames (with at least one non-matching pair in between) and the ninth or 10th frame, one was double-matching (in 75% of trials), which was always the seventh frame, and the others were non-matching. Altogether, there were 12 trials of 10 pairs. For the purpose of design, the 12 trials were split into three sub-blocks of four trials. In each sub-block, one trial had only colour-single-matching pairs, one trial had only shape-single-matching pairs, and two trials had a mixture of colour- and shape-single-matching pairs. Additionally, one trial required only left responses, one trial required only right responses, and two trials required a mixture of left and right responses. A double-matching pair was present in three of the four trials in each sub-block. Replication of the same stimulus pairs was restricted and did not occur within trials or sub-blocks.

The task required identification of pairs of items that shared the same colour or the same shape. If an item pair matched on either dimension, the participant was to respond by pressing one of two response keys (with their right hand), contingent upon the position of the tick (left or right). Participants were then informed that, towards the end of each trial, they might see an item pair that matched in both colour *and* shape. The RI Present group was instructed to ignore these double-matching frames whereas the RI Absent group was instructed to say "double" out loud in addition to pressing the appropriate response key. All key presses were recorded in PsyScope and were attributed to a frame if they occurred within 2800 ms (two complete frames plus intervals) of its onset. The participants were not informed of this time limit but simply that they were to respond as quickly but as accurately as possible. Participants initially went through an example trial on paper with the experimenter and then completed a practice trial. They were then asked to repeat the rules and to answer a series of questions (e.g., "what do you do if both items of a pair share both colour and shape?"). If the rules were recalled or the questions answered incorrectly, the rules were repeated and the practice trial re-run. After completion of the task, participants were asked to repeat the task rules.

2.1.3. Design

Performance measures for the colour-shape match task are presented in Table 1. Key measures on the task were criterion fails, which took into account performance across task elements and were considered a measure of overall success on a trial, critical error (RI Present), which was the 'no-go' measure due to its reliance on response inhibition, and misses, which was the 'go' measure. The other measures presented in Table 1 have been included to illustrate performance across the full range of task components, but our primary focus here is on critical errors and criterion fails. All measures were recorded as error as a proportion of all possible instances, except for false positivesv which were recorded as raw values. All analyses were based on the full dataset (i.e., all 12 trials rather than separately by sub-block). Culture Fair raw error scores were correlated with each performance measure separately for the RI Absent and RI Present groups. Performance scores and correlations between those scores and Culture Fair error were compared across groups. The specific sequencing of trials varied pseudo randomly and was maintained for all participants in order to hold the level of response prepotency constant.

2.2. Results

Table 2 presents performance scores (upper panel) and correlations with g (as estimated by Culture Fair; lower panel). Analysis of covariance (ANCOVA; controlling for Culture Fair error) revealed that performance was significantly worse in the RI Present group relative to the RI Absent group for critical errors (response to double-matching 'no-go' pairs [RI Present] vs. missed response to double-matching pair [RI Absent]; $p < 0.001$), criterion fails (overall success on a trial; $p = 0.032$) and hand errors (response to the incorrect side of single-matching pair; $p = 0.038$). There was no significant difference across RI Present and RI Absent groups for misses (missed response to single-matching 'go' pair), false positives (response to non-matching 'neutral' pair), or reaction time ($p \geq 0.1$ for all comparisons). Although there was a general trend for reduced error and faster response time as participants progressed through the task, once a requirement had been satisfied participants commonly produced error on that requirement on later trials – and this was particularly the case for the critical double-matching frames requiring no response. Twenty-nine of the 38 participants in the RI Present group (76%) produced a critical error in trial 1, with this figure reducing to 19 (50%) on the final double-matching frame (trial 11).

Table 2
Performance scores (A, upper panel) and correlation with g (B, lower panel) across RI Present and RI Absent groups in Experiment 1.

A	RI Present ($n = 38$)		RI Absent ($n = 31$)		Comparison across groups		
	Performance				ANCOVA		
	M	SD	M	SD	$F(1, 66)$	p	η_p^2
Critical error	0.65	0.25	0.34	0.34	18.51	<0.001	0.219
Criterion fail	0.66	0.20	0.53	0.29	4.82	0.032	0.068
Miss	0.26	0.15	0.20	0.15	2.16	0.146	0.032
Hand error	0.08	0.08	0.12	0.11	4.46	0.038	0.063
False positivek	0.08	0.18	0.02	0.03	2.81	0.098	0.041
False positivev	–	–	2.71	1.75	–	–	–
Reaction time (ms)	1201	187	1168	222	0.21	0.648	0.003

B	Correlation with g				Fisher's z-test	
	$r(36)$	p	$r(29)$	p	$z(67)$	p
Critical error	0.56	<0.001	0.00	0.999	2.50	0.012
Criterion fail	0.76	<0.001	0.27	0.142	2.84	0.005
Miss	0.52	<0.001	0.38	0.035	0.70	0.484
Hand error	0.40	0.013	0.27	0.142	0.58	0.562
False positivek	0.48	0.002	0.05	0.789	1.87	0.062
False positivev	–		0.27	0.142	–	–
Reaction time (ms)	0.24	0.147	0.48	0.006	1.10	0.271

The majority of performance measures (all but reaction time) were significantly correlated with *g* (as estimated by Culture Fair raw error scores) in the RI Present group, but only the miss–*g* and reaction time–*g* correlations reached significance in the RI Absent group. Fisher's z-test of difference for independent correlations revealed that the correlation with *g* was significantly stronger in the RI Present group relative to the RI Absent group for critical errors, $z(67) = 2.50$, $p = 0.012$, with *r*(RI Absent) outside the 95% confidence interval (CI) [0.27, 0.75] for *r*(RI Present), calculated using Fisher's z transform. The criterion fail–*g* correlation was also significantly stronger in the RI Present group, $z(67) = 2.84$, $p = 0.005$, with *r*(RI Absent) outside the 95% CI [0.58, 0.87] for *r*(RI Present). There was no significant difference for misses, hand errors, false positives, or reaction time ($p > 0.2$ in all comparisons except hand errors which was marginal at $p = 0.062$). Fig. 2 demonstrates these patterns of correlation, with more than twice as many critical errors and criterion fails in the RI Present version of the task produced by participants > 1 SD below the sample mean on the Culture Fair in comparison to those >1 SD above the mean (upper panel). The absence of a meaningful relationship between these performance measures and Culture Fair in the RI Absent task variant is demonstrated in the lower panel, although the two participants > 1 SD above the Culture Fair mean performed disproportionately better than the other participants in this group. All participants correctly repeated the instructions at the end of testing.

2.3. Discussion

The inhibition (RI Present) task was clearly highly demanding. It was expected that difficulties in withholding a response to the double matches would be observed, particularly in low *g* participants, but that high *g* participants would be able to satisfy this requirement. In fact, virtually all participants were unable to prevent a response to these stimuli. Nevertheless, *remembering* the task instructions was not difficult; there were a few, well-specified rules and no obvious element of problem-solving. All the participants correctly repeated the task rules once the experiment had ended, demonstrating that they did not forget what was required of them. Moreover, participants typically reacted verbally immediately upon producing a response to the double matching frames, indicating that the demand to inhibit the response was in mind but the strength of two response tendencies (i.e., colour match and shape match) was such that the instruction was overturned and the key inadvertently pressed. Under the considerable task demands the observed behaviour emulates that sometimes described in patients with frontal lobe damage: production of erroneous behaviour despite simultaneous correct verbalisation of task rules (Luria, 1966, 1973).

Performance on critical double-matching items was significantly correlated with *g* when a strong response inhibition demand was required, but the relationship was negligible when replaced by an alternative positive response. Additionally, performance across all task constraints was more closely correlated with *g* when the task encompassed the response inhibition demand, indicating that difficulties in resolving response conflicts are particularly common in individuals at the lower end of the *g* distribution. The findings raise the prospect that a requirement to inhibit prepotent response tendencies may be a fundamental "risk factor" for the recruitment of *g* in task performance. Nevertheless, because, on average, performance was better on the RI Absent task, the difference in *g*-correlations across tasks may be driven by broader issues associated with differences in overall task difficulty. We address this issue in Experiment 3, following a replication of our RI Present task in a younger group of participants.

3. Experiment 2: a replication in younger participants

Experiment 1 was conducted with older participants that, on average, scored 0.5 SDs below published adult Culture Fair norms, and there was a comparative absence of those scoring at the high end of the normative distribution. Variability in performance is known to increase with age (Rabbitt, 1993), desirable to the extent that a broad range of scores may be collected. A negative consequence, however, is that different cognitive abilities may show differential sensitivities to age (e.g., Kievit et al., 2014) and it cannot be assumed that similar patterns of performance will be observed across different age groups. In Experiment 2, we therefore assessed the sensitivity of performance on the response inhibition task to Spearman's *g* in a younger group of participants. It was predicted that the level of error would be lower than that observed in Experiment 1, but performance would remain highly correlated with Culture Fair scores.

3.1. Method

3.1.1. Participants

Thirty neurologically healthy adults (7 males, 23 females), aged between 30 and 50 years ($M = 40.7$, $SD = 6.1$), were recruited from the MRC Cognition and Brain Sciences Unit panel of volunteers. Culture Fair error ($M = 12.43$ $SD = 5.27$) was significantly lower than in Experiment 1 (RI Present), a difference of approximately 10 IQ points, using normalised standard scores (Cattell & Cattell, 1973). All participants gave informed consent and were paid to participate.

3.1.2. Tasks and procedure

The task instructions, procedure and performance measures were identical to those administered to the RI Present group in Experiment 1.

3.2. Results

Performance was significantly better in comparison to Experiment 1 on most measures—including the key measures, that is, misses (missed response to single-matching 'go' pair), critical errors (inappropriate response to double-matching 'no-go' pair) and criterion fails (overall success on a trial)—although reaction times were closely similar (Table 3, upper panel). However, when controlling for Culture Fair error, there was a convergence of performance across experiments such that no comparisons reached statistical significance (at $p = 0.05$). Thus, the performance differences between the younger (Experiment 2) and older (Experiment 1) groups appear largely attributable to fluid intelligence (as measured by Culture Fair). The proportion of participants producing a critical error on trial 1 (73%) was similar to that observed in Experiment 1 (76%), although there was a steeper practice effect across trials (only 13% of participants produced a critical error on the final double matching frame, compared to 50% of participants in Experiment 1).

Correlations of performance against Culture Fair were numerically smaller than those observed in Experiment 1 (Table 3, lower panel). Nevertheless, direct comparison of coefficients across experiments (using Fisher's z-test of difference for independent correlations) revealed no statistically significant differences in the strength of these correlations. The correlations were significant for critical error and criterion fails. Fig. 3 presents a comparison of performance in younger (Experiment 2) against older (Experiment 1) participants, grouped by Culture Fair z score bins of width 0.5 SD. Despite better Culture Fair performance (with a complete absence of participants in the lowest bin), the pattern of performance in younger participants was similar to the older group, with those scoring above the Culture Fair mean producing disproportionately fewer critical errors and criterion fails than those below the mean.

3.3. Discussion

Results from Experiment 2 indicate that the sensitivity of the inhibition (RI Present) task to variations in Culture Fair performance is robust and operates in younger, as well as older, participants. Close correspondence was observed across experiments in the proportion of critical errors and criterion fails produced when performance was pooled by

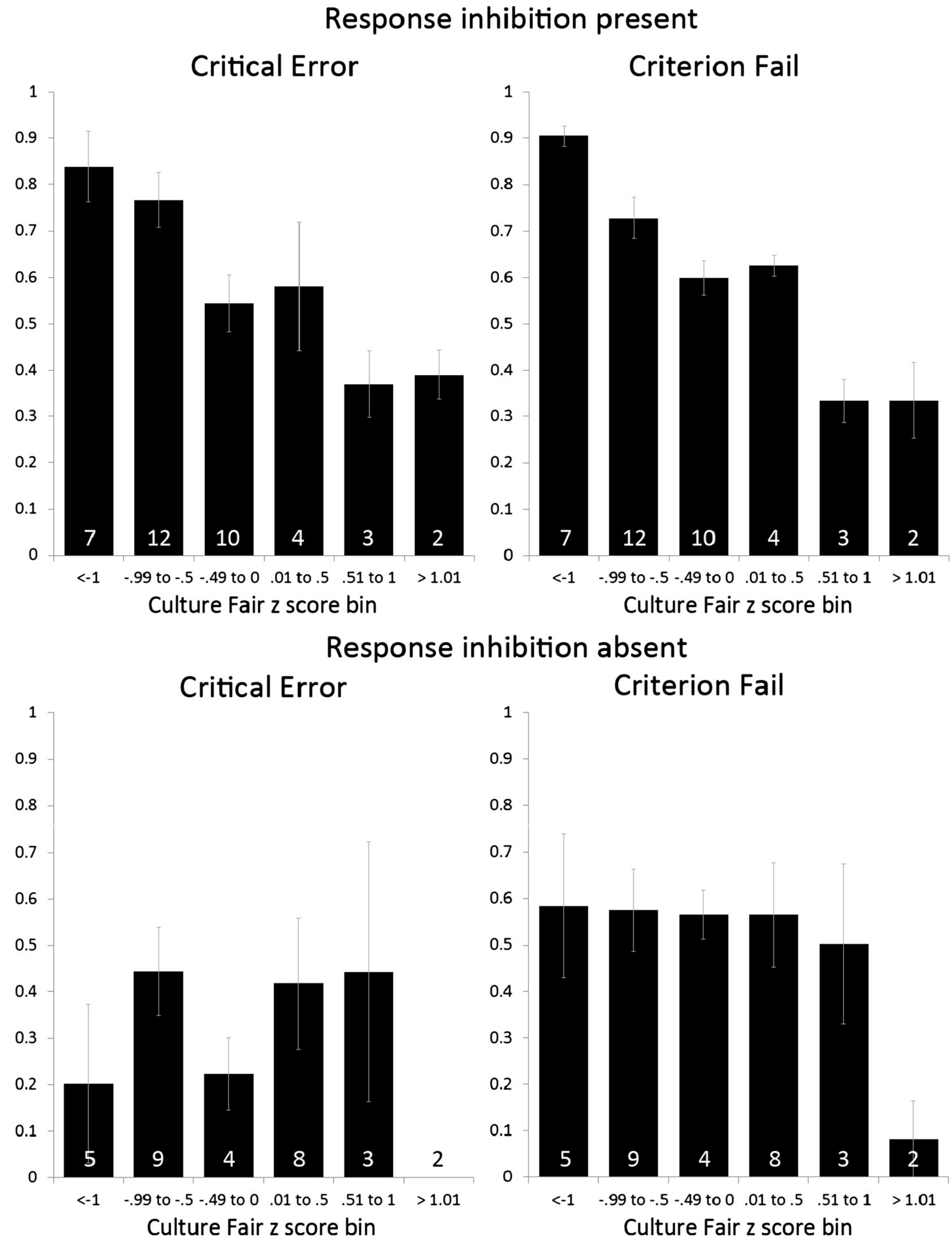

Fig. 2. Critical error and criterion fail in the RI Present and RI Absent groups across Culture Fair z score bins in Experiment 1. Error bars represent standard error for data within the respective bin. The numbers in the individual columns represent number of participants in that z score bin.

Culture Fair score. Indeed, once Culture Fair performance was partialled out, there were no reliable differences in level of performance between the younger and older groups.

The findings from Experiments 1 and 2 indicate that difficulties in resolving response conflict operate in the normal population, and are particularly common in individuals at the lower end of the *g* distribution. However, using the criterion of errors produced, the RI Present task was clearly more demanding than the RI Absent task, and this raises the possibility that the strong performance–*g* correlations were driven by more general demands associated with increased task difficulty or complexity rather than by the demand for response inhibition. This issue is addressed in Experiment 3.

4. Experiment 3: response inhibition versus task complexity

In Experiment 3 we focused on real-time processing demands, as we did in Experiments 1 and 2, examining the role of task complexity in the context of response inhibition. From a limited attentional capacity perspective, selectively attending to one or more task demand will reduce the capacity for attending to additional demands (e.g., Broadbent, 1958; Desimone & Duncan, 1995). Consequently, the more task constraints there are, the more likely it is that one of these will fail to be selected or acted upon. In Experiment 3, therefore, we examine whether the association of task performance with *g*, demonstrated in Experiments 1 and 2, can be explained primarily on the basis of

Table 3
Performance scores (A, upper panel) and correlation with *g* (B, lower panel) across younger (Experiment 2) and older (Experiment 1) groups.

A	Younger ($n = 30$)		Older ($n = 38$)		Comparison across groups					
	Performance				ANOVA			ANCOVA		
	M	*SD*	*M*	*SD*	$F(1, 66)$	*p*	η_p^2	$F(1, 65)$	*p*	η_p^2
Critical error	0.44	0.30	0.65	0.25	9.80	0.003	0.129	1.41	0.239	0.021
Criterion fail	0.46	0.20	0.66	0.20	13.99	<0.001	0.175	1.88	0.175	0.028
Miss	0.13	0.14	0.26	0.15	12.86	0.001	0.163	3.24	0.077	0.047
Hand error	0.03	0.04	0.08	0.08	8.42	0.005	0.113	1.91	0.172	0.029
False positivek	0.02	0.01	0.08	0.18	3.03	0.086	0.044	0.00	0.950	0.000
Reaction time (ms)	1180	140	1201	187	0.27	0.605	0.004	0.22	0.641	0.003
B	Correlation with *g*				Fisher's z-test					
	$r(28)$	*p*	$r(36)$	*p*	z(66)			*p*		
Critical error	0.46	0.014	0.56	<0.001	0.53			0.596		
Criterion fail	0.54	0.003	0.76	<0.001	1.72			0.126		
Miss	0.32	0.097	0.52	<0.001	0.96			0.337		
Hand error	0.23	0.239	0.40	0.013	0.74			0.459		
False positivek	0.18	0.359	0.48	0.002	1.33			0.184		
Reaction time (ms)	0.32	0.097	0.24	0.147	0.34			0.734		

task complexity (measured by the number of task constraints) or whether there is something fundamental about response inhibition in the performance-*g* relationship. In order to address this issue we manipulated the tasks such that the complexity of the RI Absent task was increased and the complexity of the RI Present task was reduced. In the Simple RI Present task, the requirement to press left or right was removed, leaving only one response key for single-matching frames. In the Complex RI Absent task, items varied along three dimensions: colour, shape, and shading. We predicted that although the Simple RI Present task would be easier to perform than the RI Present task of Experiments 1 and 2, sensitivity to variations in fluid intelligence would remain. We also predicted that moderate correlations between performance and *g* would be found in the Complex RI Absent task, due to increased attentional demands, but that these associations would not exceed those observed in Simple RI Present, despite the likelihood of higher overall level of error. This result would suggest that the role of *g* in task performance increases with task complexity but that a process of action control via suppression of inappropriate response tendencies may be a fundamental risk factor for the recruitment of *g* that operates relatively independently of task complexity.

4.1. Method

4.1.1. Participants

Participants were 36 neurologically healthy adults (7 males, 29 females) aged between 22 and 52 years ($M = 38.5$, $SD = 9.5$); mean Culture Fair error was 11.92 ($SD = 4.87$). All participants were recruited from the MRC Cognition and Brain Sciences Unit paid panel of volunteers and each participant gave informed consent.

4.1.2. Tasks and procedure

4.1.2.1. Simple RI Present colour-shape match task. The task and instructions were identical to those administered to the RI Present group of Experiment 1 with the exception that the ticks and crosses were removed from the centre of the objects and that participants were required to press one button for all single-matching pairs (see Fig. 4, left panel).

4.1.2.2. Complex RI Absent colour-shape match task. The task and instructions were identical to those administered to the RI Absent group of Experiment 1 with the following exceptions: a) the ticks and crosses in the centre of the objects were replaced by a horizontal line which indicated

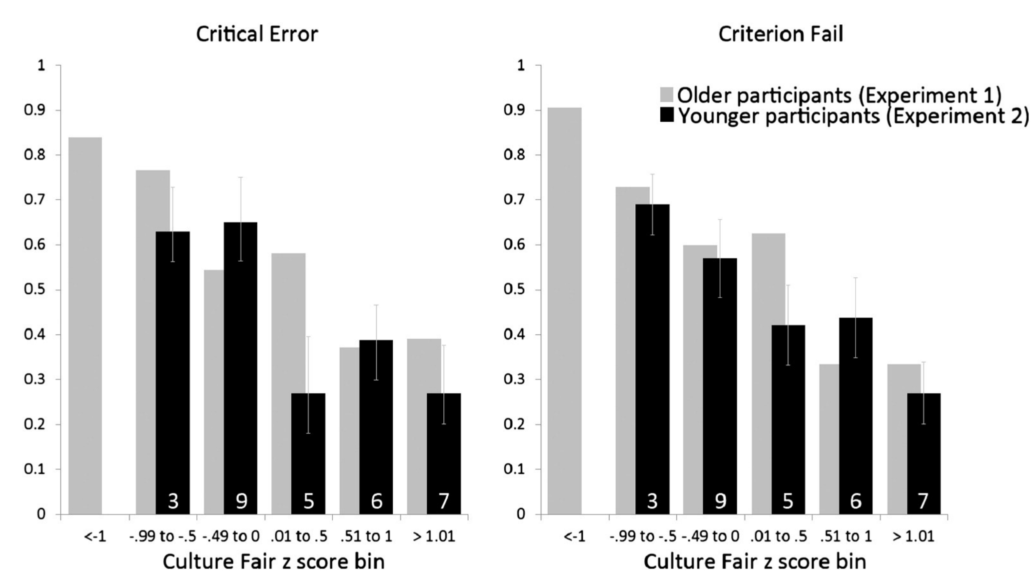

Fig. 3. Critical error and criterion fail in the younger (Experiment 2) and older (Experiment 1) groups across Culture Fair z score bins. Charts show RI Present data only. Error bars represent standard error. The numbers in the individual columns represent the number of participants in that z score bin.

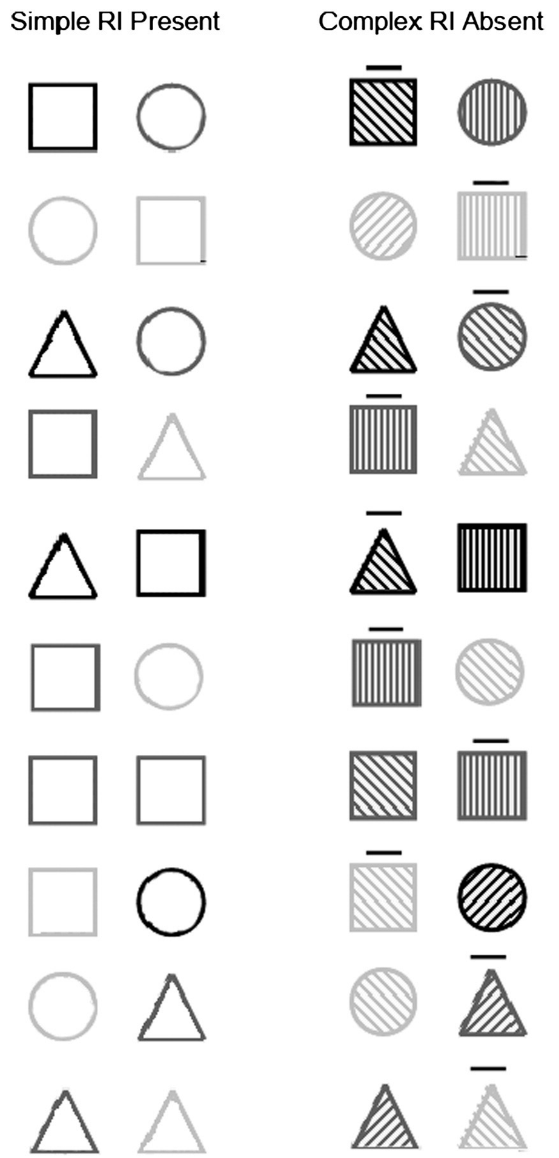

Fig. 4. Representation of typical Simple RI Present and Complex RI Absent trials in the colour-shape match task. Black, mid grey, and light grey objects are presented in blue, red, and green, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the required response (left or right) with the right hand as shown in Fig. 4 (right panel); b) pairs varied across three levels of shading (as well as across colour and shape). In 75% of trials one pair matched in shade, which also required a response (with the left hand) according to the position of the horizontal line. A shade match occurred only in frames that did not match in colour or shape. Thus, participants were required to press the appropriate key (left or right) when items matched on colour and/or shape and when items matched on shade. For the 75% of trials in which the seventh frame matched on both colour *and* shape, participants were required to press the relevant key and say “double” (as was the case for the RI Absent group in Experiment 1).

Participants completed the two versions of the colour-shape match task—Simple RI Present and Complex RI Absent—with the Culture Fair (Scale 2, Form A) being administered in between the two task variants. Instructions were given in the same manner as in Experiment 1, with the removal/addition of performance requirements to reflect the demands of these task variants. For each task, participants initially worked through an example trial on paper with the experimenter and completed a practice trial, and were then asked to repeat the rules. If recalled incorrectly, the rules were repeated and the practice trial was re-run. The experimental trials, therefore, did not proceed until clear understanding of the instructions had been demonstrated. Participants were also asked to recall the instructions after completion of each task.

4.1.3. Design

Performance measures were the same as those in Experiment 1, with the addition, in the Complex RI Absent task, of *misssh* (missed response to single-matching items matching in shade), *between hand error* (response with the wrong hand), *neglected requirements* (complete failure to attend to one or more requirements throughout the course of the task), and *perfect trials* (satisfaction of all task demands on individual trials). False positivesv and neglected requirements were recorded as raw values whereas the remainder of performance measures were recorded as proportion of error. Performance scores and correlations between performance scores and Culture Fair error were compared across the Simple RI Present and Complex RI Absent tasks. Ordering of the two experimental tasks was counterbalanced, with half of the participants completing the Simple RI Present task first.

4.2. Results

Performance scores (upper panel) and correlations with *g* (lower panel) are presented in Table 4. On the Simple RI Present task, error on all measures was low, with participants on average making fewer than one miss (missed response to single-matching ‘go’ pair), critical error (inappropriate response to double-matching ‘no-go’ pair), or false positive (response to non-matching ‘neutral’ pair) across all 12 trials. However, over 25% of participants failed to suppress a response to the double-matching frame in trial 1. Overall, a mean of 11% of trials were failed. On the Complex RI Absent task, though, a substantial proportion of error was found on all measures. Participants failed to respond verbally to over 50% of double-matching items (critical errors)

Table 4
Performance scores (A, upper panel) and correlation with *g* (B, lower panel) for Simple RI Present and Complex RI Absent tasks in Experiment 3.

A	Simple RI Present		Complex RI Absent	
	Performance			
	M	*SD*	*M*	*SD*
Critical error	0.08	0.07	0.53	0.30
Critical error (trial 1)	0.28	0.45	–	–
Criterion Fail	0.11	0.12	–	–
Misscs	0.02	0.04	0.17	0.15
Misssh	–	–	0.70	0.25
Within hand error	–	–	0.14	0.13
Between hand error	–	–	0.10	0.13
False positivesk	0.004	0.01	0.03	0.11
False positivesv	–	–	2.14	4.86
Perfect trials	–	–	0.09	0.18
Neglected requirements	–	–	0.28	0.57
Reaction time (ms)	844	96	1258	175
B	Correlation with *g*			
	r(34)	*p*	*r*(34)	*p*
Critical error	0.44	0.007	0.40	0.016
Critical error (trial 1)	0.42	0.011	–	
Criterion Fail	0.24	0.159	–	
Misscs	0.13	0.450	0.13	0.450
Misssh	–		0.47	0.004
Within hand error	–		0.01	0.954
Between hand error	–		0.08	0.643
False positivesk	0.02	0.908	−0.23	0.177
False positivesv	–		−0.08	0.643
Neglected requirements	–		0.38	0.022
Reaction time (ms)	0.34	0.042	0.10	0.562

Note. Missch = no response to colour or shape single-matching (GO) pairs; Misssh = no response to shade single-matching pairs; False positivesk = response (key press) to non-matching pairs; False positivesv = response (verbal) to single or non-matching pairs; $n = 36$.

and missed over two-thirds of items matching on shade (*misssh*). Just over one of the 12 trials was completed without error (perfect trials) and a minority of participants showed a complete failure to attend to one or more requirements throughout the course of the experiment (neglected requirements). Nearly 20% of single-matching colour/shape items were missed. Of the targets that were responded to, an incorrect key was commonly pressed (14% with the correct hand and 10% with the incorrect hand). False positives occurred infrequently, both in terms of erroneous key responses and verbal responses to non-matching items.

Correlations with *g* were typically lower when the RI Present task was simplified, but remained highly significant for critical errors despite better performance on this measure (and all other measures). Significant correlations emerged in the Complex RI Absent task. Fig. 5 demonstrates the trend for better performance in participants at the higher end of the *g* distribution. On the Simple RI Present task, all 10 participants scoring >1 SD above the mean on the Culture Fair were able to withhold responding to double-matching items on the first trial, whereas of those scoring >1 SD below the population mean on the Culture Fair, over 50% produced a critical error. The chart of criterion fails also clearly indicates a pattern of improved performance with increasing Culture Fair score; that this correlation did not reach significance appears primarily due to performance in those participants in the lowest z score bin (a correlation of $r = 0.37$, $p = 0.044$ was observed when omitting those participants from the analysis). On the Complex RI Absent task, neglected requirements were largely restricted to participants at the low end of the *g* distribution. Other significant correlations also emerged, including for perfect trials, as indicated in Table 4. Instructions were recalled correctly at the end of testing.

4.3. Discussion

Experiment 3 demonstrates that complexity, or the number of concurrent task demands, plays an important part in the involvement of Spearman's *g* in task performance. The Simple RI Present task was associated with more moderate correlations with *g* than was found with the standard RI Present task employed for Experiment 1, although the correlations were more consistent with Experiment 2. This result implies that, while the reversal demand may represent an important part of that association, as indicated by the sensitivity of critical errors to Culture Fair performance, other factors may have a mediating influence. The reintroduction of a relationship between error and Culture Fair in the Complex RI Absent task suggests that the number of requirements that must be concurrently attended to partly determines the strength of the association. The observation that elements of performance on an easier version of the RI Present task predicted Culture Fair scores suggests that a process of action control via management of response potency may be a basic process underlying *g*. However, involvement of *g* in this type of response suppression is complicated by more general issues of task complexity.

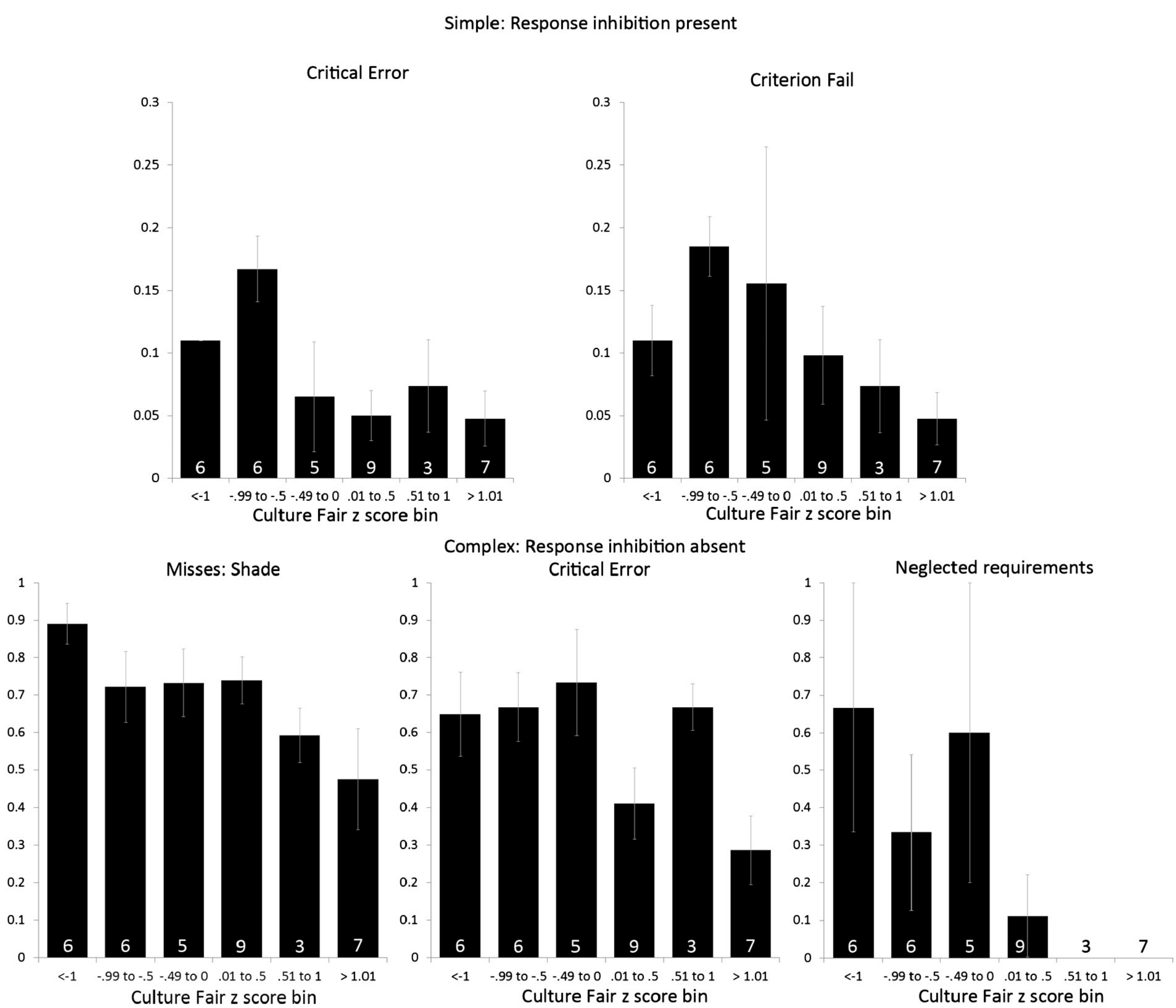

Fig. 5. Critical error and criterion fail in the simple RI Present task (upper panel) and misses (shade), critical error and neglected requirements in the complex RI Absent task across Culture Fair z score bins in Experiment 3. Error bars represent standard error. The numbers in the individual columns represent the number of participants in that z score bin.

In the Complex RI Absent task, error was most readily observed on the shade requirement, which was specified last. That the order of instructions is important in predicting which demand will be failed cannot be stated with any certainty, as all participants received the same order; the shade demand may simply be more difficult to satisfy than the others. However, the finding is consistent with the suggestion of a primacy effect, whereby requirements specified early are the most likely to be satisfied (Duncan et al., 1996). Of the various task requirements, this was also the one most sensitive to variations in *g* (although not significantly so). This finding is consistent with a reduced WMC explanation for individual differences in performance.

Items matching on colour and shape (in both tasks) and on shade (in Complex RI Absent) drove the *g*-correlations observed in this experiment, suggesting that frequency of occurrence may also be important in determining relative success on demanding tasks. The requirement to respond to frames matching on colour or shape was reinforced more frequently than other demands, therefore affecting a response bias towards these frames. In every trial, there were three frames matching on colour or shape, with a maximum of one shade match and one colour and shape match per trial. It is intuitive that a task demand is less likely to lose control of behaviour if it is being regularly reinforced by stimuli matching that demand. Our findings indicate that variability in the capacity for effectively counteracting a response bias in line with task goals is associated with individual variability in *g*.

5. Experiment 4: response inhibition versus task instruction complexity

In Experiment 4 we focus on the complexity of task instructions to address recent claims that the total complexity of task instructions (the "task model"), rather than real-time task execution demands, are most closely associated with measures of Spearman's *g* (Dumontheil et al., 2010; Duncan et al., 2008; Duncan et al., 2012; Roberts & Anderson, 2014). The claims were based on a series of tasks in which complexity was manipulated by varying the task instructions, which could be presented as a full or a reduced set (Duncan et al. 2008). In the full instructions condition requirements for two tasks were explained, and the participants were told that information for one of the tasks could be temporarily discarded. In the reduced instructions conditions, requirements for one task were explained, and after completion, requirements for the second task were explained. We have applied a different approach in which we manipulated WM load by splitting the same executable task information into two or four distinct chunks of information.

Participants completed the RI Present task of Experiments 1 and 2, and an RI Absent task which was operationally identical to the RI Present task except that double-matching pairs were never actually presented. One group of participants—the Two Rule group—were given task instructions chunked into two rules (assumed to represent an efficient mental representation of the task), and another group—the Four Rule group-–were given the same instructions instead chunked into four rules (assumed to represent an inefficient task representation). This approach provided the opportunity for investigating a) the importance of response inhibition in the context of efficiently and inefficiently constructed task instructions (while holding the actual operative task related information constant), and b) the effects of modelling task requirements which are not required during actual task execution. We predicted that performance–*g* correlations would be stronger for four-rule relative to two-rule instructions, despite holding constant information directly related to the task as well as the task execution demands themselves. If efficiency of task modelling, rather than execution, is fundamental to Spearman's *g* we might also expect to find evidence that reassembly of task rules, towards greater efficiency, would be restricted to those participants at the higher end of the *g* distribution. If performance–*g* correlations remain strong across both RI Present and RI Absent tasks in the four-rule condition, this would be most consistent with the claim that task modelling rather than task execution is the more fundamental risk factor for the recruitment of *g*.

5.1. Method

5.1.1. Participants

One-hundred neurologically healthy adults (37 males, 63 females), aged between 18 and 63 years, were recruited from the Department of Psychology, Anglia Ruskin University in Cambridge and the wider local community. Psychology undergraduates were recruited via an on-line study recruitment system which granted course credit; community volunteers were recruited via word-of-mouth and did not receive any payment. None of the participants were carried over from previous experiments. The Two Rule group comprised 50 adults (20 males, 30 females) aged between 18 and 62 years ($M = 30.98$, $SD = 12.56$). The Four Rule group comprised 50 adults (17 males, 33 females) aged between 18 and 63 years ($M = 29.22$, $SD = 12.78$). Culture Fair error was equivalent across the Two Rule ($M = 9.98$, $SD = 4.09$) and Four Rule ($M = 11.54$, $SD = 5.12$) groups, $t(98) = 1.68$, $p = 0.096$, $d = 0.339$. As in previous experiments, the sample sizes are similar to those of studies addressing similar research questions (e.g., Duncan et al. 2008; Roberts & Anderson, 2014).

5.1.2. Tasks and procedure

Participants completed the Culture Fair (Scale 2, Form A) followed by the two versions of the colour-shape match task in succession. The RI Present task variant was identical to the 12 trials of Experiment 1 (see Fig. 1); the RI Absent task variant was identical to RI Present except the nine double-matching frames were replaced by nine additional non-matching frames. The instructions of previous experiments were modified here. Participants were informed that pairs of coloured objects, one containing a tick and the other containing a cross, would be presented one at a time in quick succession. They were informed that objects could share either the same colour or the same shape and that towards the end of each sequence, they might see a pair that shares both the same colour and the same shape. Half of participants were then given two rules to follow and the other half of participants were given four rules to follow; essentially the information contained across rule formats was the same, but was presented in the form of either two or four chunks (see Table 5). Participants worked through an example trial with the experimenter as in previous experiments but here there was no practice trial. The rules were then repeated for a second time and the participant was asked to repeat the rules (which were recorded). Unlike previous experiments, participants were not asked a series of questions about the task. If parts of the rules were omitted, the experimenter repeated the appropriate rule until the participant was able to repeat all parts of all rules. This ensured that task instructions were effectively stored in short-term memory such that every participant knew what to do. Immediately before each task, participants were told that the presentation of items would be very fast, to not be surprised at this, to just do their best, and that they would get a chance for a short break after every 10th pair of items. Immediately after each

Table 5
Task instruction rules for Experiment 4.

Group	Rule	Description
Two Rule	Rule 1	Respond to items that match in colour OR in shape by pressing the side corresponding to the tick
	Rule 2	Ignore items that do not match in colour or shape, and items that match in both colour AND shape
Four Rule	Rule 1	Respond to items that match in colour by pressing the side corresponding to the tick
	Rule 2	Respond to items that match in shape by pressing the side corresponding to the tick
	Rule 3	Ignore items that do not match in colour or shape
	Rule 4	Ignore items that match in both colour AND shape

version of the task, participants were asked to repeat the rules as they remembered them. Key presses were attributed to a frame if they occurred within 1200 ms of its onset.

5.1.3. Design

Performance measures were again identical to Experiment 1 (RI Present). All performance measures were recorded as proportion of error. Performance scores and correlations between performance scores and Culture Fair error were compared across RI Absent and RI Present task variants and across the Two Rule and Four Rule groups. The order of task variants was counterbalanced across participants, with half of participants in each group completing the RI Present task first.

5.2. Results

Performance scores were strikingly similar across the Two Rule and Four Rule groups, as displayed in Table 6; ANCOVA showed that the only performance measure that differed across rule groups was false positives in RI Present (response to non-matching 'neutral' pair; $p = 0.04$). Paired t-tests (see Table 7) confirmed that, in each rule group, error was significantly greater for critical errors than for any other measure (excluding criterion fails, i.e., overall success on a trial; all $p < 0.001$, significant using the Bonferroni-corrected alpha level of $p < 0.05/6 = 0.008$). There was a trend for greater error and slower reaction times in RI Present compared to RI Absent for every performance measure; in each rule group, this effect was significant for misses (missed response to single-matching 'go' pair), reaction time, and criterion fails (when both including and excluding critical error as a criterion for trial failure; all $p < 0.001$, significant using the Bonferroni-corrected alpha level of $p < 0.05/6 = 0.008$; see Table 7). The proportion of participants failing to withhold a response to critical double-matching frames (critical error) decreased from 32 to 15% (in Two Rule) and from 54 to 32% (in Four Rule) across the 12 trials. Again, successful inhibition did not prevent later error on critical items; this was true for 78% of the Two Rule group and for 80% of the Four Rule group.

Performance–g correlations across rule groups (Two Rule vs. Four Rule) are presented in Table 8. Fisher's z-tests showed that the critical error–g correlation was stronger in the Four Rule group, with r(Two Rule) outside the 95% CI [0.31, 0.71] for r(Four Rule), although this difference was marginally outside the statistical threshold of $p = 0.05$ ($p = 0.051$, two-tailed). The criterion fail–g correlation was significantly stronger in Four Rule, with r(Two Rule) outside the 95% CI [0.45, 0.79] for r(Four Rule). There was a trend for higher performance–g correlations a) for critical errors relative to other measures, and b) in RI Present relative to RI Absent, but only in the Four Rule group. Williams-Hotelling t-tests revealed that the critical error–g correlation did not differ significantly from other performance–g correlations in Two Rule (all $p > 0.1$), but in Four Rule, the correlation with g was significantly higher for critical errors relative to false positives in RI Present, $t(47) = 2.22$, $p = 0.03$, reaction time in RI Present, $t(47) = 3.26$, $p = 0.002$, and reaction time in RI Absent, $t(47) = 3.15$, $p = 0.001$. These latter two findings, however, would be expected due to negligible correlations for response time in each block. Additionally, performance–g correlations did not differ significantly across RI Present and RI Absent task variants for any measure in the Two Rule group (all $p > 0.1$), but the RI Present criterion fail–g correlation was significantly stronger in RI Present relative to RI Absent in the Four Rule group, $t(47) = 2.32$, $p = 0.02$.

Fig. 6 displays performance scores for critical errors and criterion fails across Culture Fair z-score bins. All charts demonstrate the stronger relationship between task performance and g in four-rule relative to two-rule conditions, and in RI Present relative to RI Absent task variants. For example, there was a 20% difference in critical error between lower and higher g participants (those scoring > 1 SD below and above the mean, respectively) in Two Rule, compared to a 34% difference in Four

Table 7
Paired t-tests comparing performance scores between RI Present and RI Absent tasks (A, upper panel) and between critical errors and other measures (B, lower panel) in Experiment 4.

A	Two Rule			Four Rule		
	RI Present vs. RI Absent					
	$t(49)$	p	Cohen's d	$t(49)$	p	Cohen's d
Criterion fail	9.75	<0.001	1.369	10.48	< 0.001	1.669
w/o critical error	4.17	<0.001	0.560	4.33	< 0.001	0.682
Miss	3.87	<0.001	0.503	4.68	<0.001	0.743
Hand error	0.92	0.362	0.238	0.32	0.750	0.177
False positivek	1.59	0.118	0.223	0.49	0.626	0.127
Reaction time (ms)	3.90	<0.001	0.563	5.32	< 0.001	0.757
B	Critical error vs. other measures					
	$t(49)$	p	Cohen's d	$t(49)$	p	Cohen's d
RI Present						
Miss	4.20	<0.001	0.648	5.86	<0.001	1.095
Hand error	9.55	<0.001	1.760	10.67	<0.001	2.233
False positivek	6.84	<0.001	1.026	10.71	<0.001	1.249
RI Absent						
Miss	9.66	<0.001	1.491	10.40	<0.001	1.835
Hand error	5.93	<0.001	0.708	7.94	<0.001	1.245
False positivek	8.23	<0.001	1.895	11.47	<0.001	2.231

Table 6
Performance scores across Two Rule and Four Rule groups in the RI Present (A, upper panel) and RI Absent (B, lower panel) tasks in Experiment 4.

A	Two Rule (n = 50)		Four Rule (n = 50)		Comparison across groups		
	RI Present				ANCOVA		
	M	SD	M	SD	$F(1, 97)$	p	η_p^2
Critical error	0.40	0.26	0.43	0.27	0.00	0.947	0.000
Criterion fail	0.50	0.27	0.49	0.27	1.01	0.317	0.010
w/o critical error	0.30	0.28	0.28	0.23	1.43	0.235	0.015
Miss	0.20	0.16	0.20	0.16	0.37	0.546	0.004
Hand error	0.04	0.04	0.06	0.09	0.64	0.426	0.007
False positivek	0.08	0.19	0.03	0.05	4.44	0.037	0.044
Reaction time (ms)	850	87	841	93	0.22	0.639	0.002
B	RI Absent				ANCOVA		
	M	SD	M	SD	$F(1, 97)$	p	η_p^2
Criterion fail	0.18	0.24	0.14	0.13	1.93	0.168	0.020
Miss	0.13	0.14	0.11	0.09	1.57	0.213	0.016
Hand error	0.03	0.04	0.05	0.08	1.67	0.199	0.017
False positivek	0.05	0.17	0.02	0.06	2.10	0.151	0.021
Reaction time (ms)	816	95	792	90	1.68	0.198	0.017

Table 8
Performance–g correlations across Two Rule and Four Rule groups in the RI Present (A, upper panel) and RI Absent (B, lower panel) tasks in Experiment 4.

A	Two Rule		Four Rule		Comparison across groups	
	RI Present				Fisher's z-test	
	$r(48)$	p	$r(48)$	p	$z(98)$	p
Critical error	0.20	0.164	0.54	<0.001	1.95	0.051
Criterion fail	0.30	0.034	0.65	<0.001	2.26	0.024
w/o critical error	0.22	0.125	0.54	<0.001	1.84	0.066
Miss	0.25	0.080	0.36	0.010	0.59	0.555
Hand error	0.33*	0.019	0.40	0.004	0.39	0.697
False positivek	0.14	0.332	0.17	0.238	0.15	0.881
Reaction time (ms)	−0.06	0.679	0.00	0.999	−0.29	0.772
B	RI Absent				Fisher's z-test	
	$r(48)$	p	$r(48)$	p	$z(98)$	p
Criterion fail	0.24	,093	0.37	0.008	0.70	0.484
Miss	0.24	0.093	0.21	0.143	0.15	0.881
Hand error	0.30	0.034	0.33	0.019	0.16	0.873
False positivek	0.19	0.186	0.32	0.023	0.68	0.497
Reaction time (ms)	0.05	0.730	−0.02	0.890	0.34	0.734

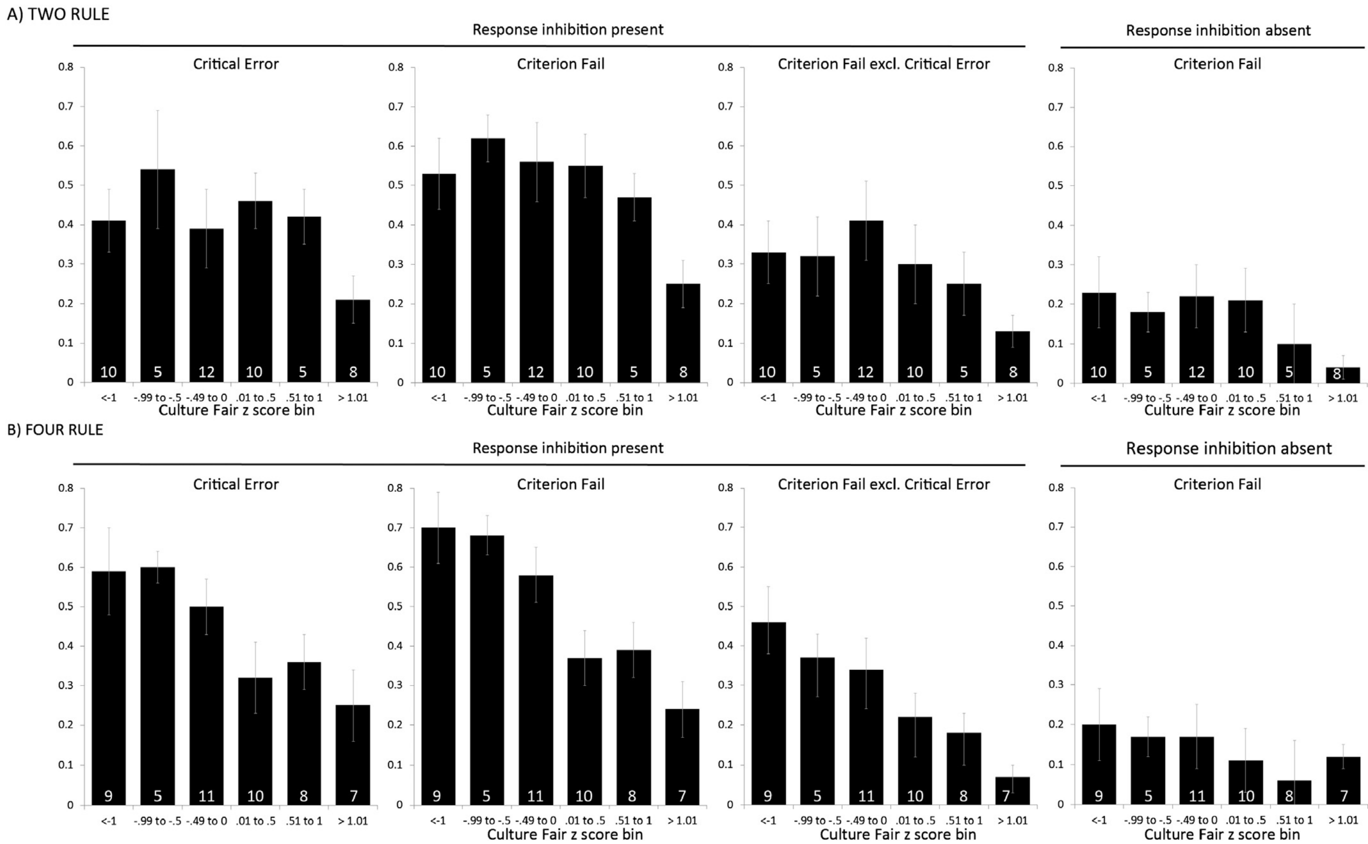

Fig. 6. Critical error and criterion fail in the Two Rule and Four Rule groups across Culture Fair z score bins in Experiment 4. Error bars represent standard error. The numbers in the individual columns represent the number of participants in that z score bin.

Rule. Criterion fails (incorporating critical error) varied from 28 to 19% (in Two Rule) and from 46 to 8% (in Four Rule) across lower and higher *g* participants, respectively, in the RI Present block. We also observed differences in the *g*-correlations between participants that completed the RI Present task variant first and participants that completed the RI Absent task first in the Four Rule group only. Fisher's z-test (two-tailed) revealed that the correlation with Culture Fair error was significantly stronger in participants (in the Four Rule group) completing the RI Absent task first relative to participants completing the RI Present task first for the following measures: critical errors, $z(48) = 2.23$, $p < 0.05$, hand errors in RI Present, $z(48) = 3.01$, $p < 0.01$, and hand errors in RI Absent, $z(48) = 2.78$, $p < 0.01$.

When asked to repeat the rules at the end of testing, participants could describe what was required of them during task execution. Some participants (Two Rule, $n = 8$; Four Rule, $n = 19$) stated the rules in a different number of chunks than specified in their administered task instructions. Although no significant differences were observed, Culture Fair error tended to be lower in those participants that reconceptualised the task towards greater efficiency (i.e., from two rules to one: $M = 7.00$ errors; or from four rules to three, two, or one: $M = 10.84$) relative to participants that did not reconceptualise the rules (Two Rule: $M = 10.17$ errors; Four Rule: $M = 10.93$).

5.3. Discussion

The results from Experiment 4 suggest that regardless of whether task instructions were administered as two or four rules, the demand to inhibit an inappropriate prepotent response tendency was relatively difficult. In both rule groups, accuracy for critical double-matching items was significantly poorer than accuracy for any other measure, with the exception of some criterion fail measures which took into account performance across task elements. This is in line with predictions, and with previous research showing that a) no-go trial performance is less accurate than go trial performance (e.g., Redick et al., 2011), and b) items specified last in instructions are more likely to be neglected than earlier specified items (Duncan et al., 1996). However, performance differed from the "goal neglect" reported in Duncan et al. (1996) in which once a requirement was satisfied it was (usually) not neglected again. Here, for the majority of participants successful inhibition of a response to no-go items did not prevent later critical error. Performance on the task that required response inhibition was also relatively difficult for participants irrespective of how the instructions were administered. Separately in each rule group, accuracy (misses) and speed (reaction time) for go items and overall performance (criterion fails) was significantly poorer in RI Present relative to RI Absent. Whether this was due to the inclusion of response inhibition demand per se, or was the consequence of RI Present having an additional task constraint, is unclear. Performance was so similar across rule groups that only one of the performance measures (false positives) differed significantly between them.

Differences between participants given task instructions as two rules and participants given task instructions as four rules did, however, emerge in the correlational analyses. In line with predictions, the correlation with Culture Fair error was significantly stronger in Four Rule relative to Two Rule for the two measures central to our hypotheses—critical errors and criterion fails in RI Present—and was numerically higher for the majority of other measures. In as much as the four rule instructions represented a more complex set of instructions, this is supportive of other work addressing the relationship between fluid intelligence and task complexity via manipulation of task instructions of a feature match task. In Duncan et al. (2008; adults) and Roberts and Anderson (2014; children) performance–*g* correlations were numerically stronger in the group administered an additional task rule even though this was actually redundant in task execution relative to the group given instructions in which this unnecessary rule was omitted. Both of these studies also reported more error, reflected in a greater amount of rule neglect, in the high-complexity instruction group. Our observation of the emergence of *statistically* stronger performance–*g* correlations in the high-complexity instruction group was, conversely, in the absence of poorer performance in this group. This could be explained by methodological differences in manipulating task instruction complexity, with both groups in our study receiving the same task executable information, albeit in different formats (two or four 'chunks'). Given this observation of equal performance in the two- and four-rule groups, the present findings cannot straightforwardly be explained with reference to task execution difficulty per se.

We also found that response inhibition was important to recruitment of *g*, but only under four-rule instructions. The critical error–*g* correlation was significant in Four Rule (and not in Two Rule, with the difference in strength between these correlations being just short of significance at $p = 0.051$), and was also shown to be significantly stronger than *g*-correlations for other performance measures (i.e., false positives in RI Present and reaction time in both task variants) only in Four Rule. The non-significant critical error–*g* correlation found in our Two Rule group supports previous work that argues against a link between response inhibition and intelligence (e.g., Friedman et al., 2006), but the significant correlation in the Four Rule group supports previous work arguing for such a relationship (e.g., Salthouse et al., 2003). This leads us to question whether conflicting results in the literature concerning the relationship between inhibition and intelligence exist, at least in part, because the complexity of task instructions, or the load imposed on the WM system by task instructions, has not been controlled. Indeed, evidence from task switching experiments suggests that switch costs occur when task instructions elicit task sets but not when task instructions suggest only stimulus-response associations (and also that when task sets are known, they dominate performance relative to simple stimulus-response mappings, Dreisbach, Goschke, & Haider, 2007). Another interesting observation was that, when asked to state task rules at the end of testing, some participants appeared to chunk task instructions into fewer or more rules than were initially specified to them at task instructions. Further, a trend for fewer Culture Fair errors was observed in those participants that stated task rules as fewer chunks, and perhaps more efficiently. It is possible, however, that because participants received no explicit instruction to impose their own order on the task rules, other instances of reconceptualisation were not identified, perhaps masking any potential significant effects.

Differences between order groups were also observed only in the Four Rule group. As well as other measures (hand errors in both task variants) the correlation with Culture Fair for items requiring response inhibition was significantly stronger for participants completing the RI Absent task first. The requirement to ignore double-matching frames was presented as a part of a rule (Rule 2) in Two Rule instructions but in isolation (Rule 4) in Four Rule instructions. If the ability to use or execute task rules is related to *g* then this finding may be explained by the greater ability of higher *g* participants to adhere to a rule that had not previously been used for correct performance. That is, by completing the 12 trials of RI Absent (i.e., a task not containing double-matching items) prior to attempting RI Present, the response inhibition requirement had not been enforced for Absent–Present participants. In contrast, Present–Absent participants use, and therefore enforce, the double-match requirement straight away. This finding suggests greater demand for operating/executing a task rule that had not been enforced during previous performance of the task. Roberts and Anderson (2014) concluded that their additional unnecessary task instruction may have increased goal neglect by disrupting the conversion of task rules into sufficiently activated goal representations. Our results suggest that real-time task experience may act to increase activation level of goal representations via enforcement of representations that are commonly used. Collectively, the findings of Experiment 4 suggest that the way in which a task is cognitively modelled is critical to the involvement of *g* in performance of that task, particularly under conditions of response conflict. Perhaps when WM capacity is stretched, additional recruitment of *g* is required to manage the execution demands of the task.

6. General discussion

Our approach to investigating individual differences in task performance has been driven in large part by neuropsychological observations of mismatch between knowledge of what is required for a given goal and the relevance of the subsequent behaviours to that goal. Dating back at least as far as Luria's (1966) ground-breaking studies of the effects of frontal lobe damage, it is clear that ability to produce goal relevant behaviour is not determined simply by the objective complexity or difficulty of task execution demands. Rather than an *inability* to perform a task effectively, there is loosening of the relationship between a given goal and the knowledge required to achieve it such that the required action(s) are simply not implemented. This phenomenon can also be observed on tasks with multiple constraints in the normal population (referred to as goal neglect; Duncan et al., 1996), and is particularly common in the performance of participants at the lower end of the *g* distribution. Our findings, like those of Duncan and colleagues, indicate that *g* confers an advantage at the level of goal conceptualisation rather than (or in addition to) task execution. Thus, it is not so much the real time processing demands of a task that drive performance–*g* correlations, but the efficiency with which the task-relevant constraints are modelled prior to, and perhaps remodelled during, actual performance.

An intuitive implication of our claim is that if the components (including response biases) inherent in a given task, however complex, are modelled in the same way by all participants, differences in performance would not be predicted in any meaningful way by a participant's relative position on the *g* distribution. Although it is impossible to determine (and control how) a task is modelled with absolute certainty, we consider this to be unlikely. In our view, variability in the mental construction/specification of complex goals is a major contributory factor distinguishing low from high *g* participants, but evidence for significant performance–*g* correlations in basic tasks with minimal execution demands (e.g., simple reaction time and inspection time paradigms; Grudnik & Kranzler, 2001; Jensen, 1998, 2006; Kranzler & Jensen, 1989; Sheppard & Vernon, 2008) suggests that factors associated with real time task execution place demands on *g* over and above recruitment associated with efficiency of task modelling. Nevertheless, in our study, as in that of Duncan et al. (2008), the body of instructions rather than the task per se appeared to drive sensitivity of performance to variations in *g*. In the current study, the status of task modelling in this relationship over more objective task characteristics is best illustrated by the observation (in Experiment 4) that the response inhibition requirement showed significantly closer sensitivity to *g* in the four-rule (relative to two-rule) task, despite a near identical level of error in the tasks. In those trials incorporating response inhibition (i.e., double-matching frames), error rates were higher on most measures of performance (irrespective of instructions) and this was particularly true for individuals who produced more errors on the Culture Fair. Although this is consistent with evidence that inhibition is a primary risk factor for recruitment of *g*, the correlational findings of Experiment 4 show that the strength of this relationship is determined in large part by the efficiency of task conceptualisation.

A recent study (Bhandari & Duncan, 2014) has provided evidence against the possibility that task instructions shape the mental representation of a novel task, as suggested by our findings. They found performance to be similar in a group given instructions chunked into distinct parts and a group given instructions in a more "interleaved" fashion. Their investigations instead suggested that goal neglect was determined by characteristics of the performed task, with the more complex sub-parts of the task (those with a greater number of task components) more likely associated with goal neglect, particularly in people at the lower end of the g distribution. They interpreted these findings as evidence for goal-directed behaviour being realised via discrete "attentional episodes" characterised by a clear, focused association between current demands of the environment and the appropriate behavioural response, suggesting that fluid intelligence may reflect the ability to convert complex requirements into "smaller more manageable parts" (Bhandari & Duncan, 2014, p. 29; see also Duncan, 2013). From such a perspective we may have expected to see better performance and lower correlation with Culture Fair scores in our four-rule condition, with the rules more distinctly mapping a specific task event relative to two-rule. It could be that the effects of instructions on task modelling are observed in tasks in which distinct task events are more easily mapped to distinct rules, such as in the feature match task employed here (as well as Duncan et al. 2008 and Roberts & Anderson, 2014; see also Dreisbach et al., 2007) but not in tasks in which distinct task parts involve a number of constraints. We argue that the contribution of *g* to cognitive performance may reflect the efficacy with which an inefficient model can be reconceptualised towards greater efficiency, with efficiency being defined as fewer distinct chunks (however complex), thereby releasing WMC for attentional reallocation (rather than splitting the task into manageable parts). Given the preliminary evidence for reconceptualisation of task information observed in our Experiment 4, we suggest that the possible link between *g* and the ability to flexibly manipulate and remodel constraints associated with complex goal-directed behaviours is likely to be a valuable focus for future investigation.

Whatever the case, we submit that models of intelligence built on correlations with objective measures of task difficulty which do not incorporate individual differences in the way in which those tasks are mentally represented are overlooking a fundamental role of intelligence: to construct action plans and realise goals through optimal allocation of resources. Participants presented with the same complex task may not, from a cognitive perspective, be undertaking the same task. High *g* confers an ability to remove redundant task information and to chunk task relevant information more efficiently. The role during execution may relate more to monitoring task content and manipulating the model in the direction of greater accuracy and efficiency. Failure to take into account potential variability in the ways in which complex tasks are conceptualised may thus limit the validity of theoretical inferences drawn on the basis of group averaging of performance scores. Further clarification of these and other possible explanatory variables driving individual differences in task performance will require the incorporation of multiple indicators and the use of confirmatory factor analysis or structural-equation modelling.

An area demanding attention in future work is the exact mechanisms by which fluid intelligence and task conceptualisation are connected, including explorations of potential alternative explanations and mediating factors. This is particularly true in light of other work providing evidence against the notion that memory and concurrent processing share a limited resource (e.g., Oberauer et al., 2004). For example, constructing and remodelling task representations could rely on processes akin to those in traditional conceptualisations of short-term memory. Our results do not support the notion that difficulties in successfully performing the experimental tasks were due to short-term memory failures in the sense of forgotten task instructions. It is nonetheless conceivable that the observed relationship between fluid intelligence and the way in which a task is cognitively modelled may dissolve when short-term storage capacity is controlled (in the same way as the relationship between fluid intelligence and WM processing factors such as attention control; e.g., Chuderski et al., 2012; Colom et al., 2008).

A recent theoretical development of the WM model distinguishes between declarative and procedural WM (Oberauer, 2009; Oberauer, Souza, Druey & Gade, 2013; see also Baddeley, 2012) with the latter serving the formation of goal-relevant stimulus-response bindings, and the bindings themselves forming the content of a higher level executive system (the *bridge*; Oberauer, 2009). The model provides a framework for quickly establishing new stimulus-response associations defined by task instructions, and a 'response focus' necessary for setting a given response apart from other possible competing responses within the task set. Thus, the model incorporates inhibitory links between response alternatives within a task set and a mechanism for excluding task irrelevant information (i.e., unless information is held in the bridge, candidate actions are unavailable for selection by the response focus). While this model requires further development, it provides an intuitively appealing theoretical framework for advancing our understanding of individual differences in cognitive ability. In our view, exploration of the relationship between task performance and the capacity/functionality of procedural WM is likely to foster important advances in this area.

In summary, the present findings indicate that the involvement of Spearman's *g* in task performance is largely determined by the relational complexity of individual task components, and that the level of effort required for optimal modelling of a task largely determines the strength of performance–*g* correlations. Performance discrepancies between high and low *g* individuals are likely to be greatest in earlier stages of novel task execution when initial modelling of task constraints may be subject to reorganisation and consolidation, with the effectiveness and time course of model stabilisation contingent upon the complexity of the task. Response conflict is an important risk factor for the engagement of *g* because it demands a clear parsing of available but incompatible responses in line with task instruction. Other constraints limit the ability to form a stable task model, with those at the lower end of the *g* distribution less able to manipulate the representation of (and relationship between) task demands in an optimal way. Future research will require systematic exploration of individual differences in the ways in which task instructions are initially modelled and then remodelled over the course of preparation for, and execution of, goal-directed complex behaviours.

Acknowledgements

Experiments 1–3 were supported by a Medical Research Council (MRC) PhD studentship awarded to PB while based at the MRC Cognition & Brain Sciences Unit. Experiment 4 was supported by an internally funded PhD studentship awarded to EC while based at Anglia Ruskin University. We wish to thank John Duncan for his encouragement and advice during the course of this work, and an anonymous reviewer for his/her helpful comments on an earlier draft.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*(4), 567–589.

Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.), *Recent Advances in Learning and Motivation* (pp. 47–90). New York: Academic Press.

Bhandari, A., & Duncan, J. (2014). Goal neglect and knowledge chunking in the construction of novel behaviour. *Cognition*, *130*, 11–30.

Bright, P. D. (1998). *The Control of Action: An Exploration of Spearman's General Factor.* Unpublished doctoral thesis Cambridge, England: University of Cambridge.

Broadbent, D. (1958). *Perception and Communication.* London: Pergamon Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Cattell, R. B., & Cattell, H. E. P. (1973). *Measuring Intelligence With the Culture Fair Tests.* Champaign, IL: Institute for Personality and Ability Testing.

Chuderski, A., Taraday, M., Nęcka, E., & Smoleń, T. (2012). Storage capacity explains fluid intelligence but executive control does not. *Intelligence*, *40*, 278–295.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: a new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments, and Computers*, *25*(2), 257–271.

Colom, R., Abad, F. J., Quiroga, M.Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606.

Colom, R., Flores-Mendoza, C., Quiroga, M.Á., & Privado, J. (2005). Working memory and general intelligence: the role of short-term storage. *Personality and Individual Differences*, *39*(5), 1005–1014.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, *30*, 163–183.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences*, *7*(12), 547–552.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

Das, J. P. (2002). A better look at intelligence. *Current Directions in Psychological Science*, *11*(1), 28–33.

Dempster, F. N., Corkill, A. J., & Jacobi, K. (1995, November). Individual differences in resistance to interference. *Paper presented at the Annual Meeting of the Psychonomic Society, Los Angeles*.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Dreisbach, G., Goschke, T., & Haider, H. (2007). The role of task rules and stimulus–response mappings in the task switching paradigm. *Psychological Research*, *71*(4), 383–392.

Dumontheil, I., Thompson, R., & Duncan, J. (2010). Assembly and use of new task rules in fronto-parietal cortex. *Journal of Cognitive Neuroscience*, *23*(1), 168–182.

Duncan, J. (2013). The structure of cognition: attentional episodes in mind and brain. *Neuron*, *80*, 35–50.

Duncan, J., Emslie, H., Williams, P., Johnson, R., & Freer, C. (1996). Intelligence and the frontal lobe: the organization of goal-directed behavior. *Cognitive Psychology*, *30*(3), 257–303.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., ... Nimmo-Smith, I. (2008). Goal neglect and Spearman's *g*: competing parts of a complex task. *Journal of Experimental Psychology: General*, *137*(1), 131–148.

Duncan, J., Schramm, M., Thompson, R., & Dumontheil, I. (2012). Task rules, working memory, and fluid intelligence. *Psychonomic Bulletin & Review*, *19*(5), 864–870.

Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 145–199). New York: Academic Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331.

Friedman, N. P., Miyake, A., Corely, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, *17*(2), 172–179.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679.

Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience*, *6*(3), 316–322.

Grudnik, J. L., & Kranzler, J. H. (2001). Meta-analysis of the relationship between intelligence and inspection time. *Intelligence*, *29*(6), 523–535.

Hasher, L. (1997). Inhibitory control over no-longer-relevant information: adult age differences. *Memory & Cognition*, *25*(3), 286–295.

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation. Vol. 22.* (pp. 193–225). San Diego, CA: Academic Press.

Hester, R., & Garavan, H. (2005). Working memory and executive function: the influence of content and load on the control of attention. *Memory & Cognition*, *33*(2), 221–233.

Jensen, A. R. (1998). *The g Factor: The Science of Mental Ability.* Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences.* Oxford: Elsevier.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of Experimental Psychology: General*, *132*(1), 47–70.

Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, & J. N. Towse (Eds.), *Variation in Working Memory* (pp. 21–48). Oxford: Oxford University Press.

Kerns, J. G., Cohen, J. D., MacDonald, A. W., III, Cho, R. Y., Stenger, A., & Carter, C. S. (2004). Anterior cingulate conflict monitoring and adjustments in control. *Science*, *303*, 1023–1026.

Kievit, R. A., Davis, S. W., Mitchell, D. J., Taylor, J. R., Duncan, J., Cam-CAN, & Henson, R. N. (2014). Distinct aspects of frontal lobe structure mediate age-related differences in fluid intelligence and multitasking. *Nature Communications*, *5*, 5658.

Kranzler, J. H., & Jensen, A. R. (1989). Inspection time and intelligence: a meta-analysis. *Intelligence*, *13*(4), 329–347.

Long, D. L., & Prat, C. S. (2002). Working memory and Stroop interference: an individual differences investigation. *Memory & Cognition*, *30*(2), 294–301.

Luria, A. R. (1966). *Higher Cortical Functions in Man.* London: Tavistock.

Luria, A. (1973). *The Working Brain.* London: Penguin.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General*, *130*(2), 199–207.

May, C. P., Hasher, L., & Kane, M. J. (1999). The role of interference in memory span. *Memory & Cognition*, *27*(5), 759–767.

Mitchell, J. P., Macrae, C. N., & Gilchrist, I. D. (2002). Working memory and the suppression of reflexive saccades. *Journal of Cognitive Neuroscience*, *14*(1), 95–103.

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions four general conclusions. *Current Directions in Psychological Science*, *21*(1), 8–14.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: a latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100.

Morse, C. K. (1993). Does variability increase with age? An archival study of cognitive measures. *Psychology & Aging*, *8*(2), 156–164.

Oberauer, K. (2009). Design for a working memory. *Psychology of Learning and Motivation*, *51*, 45–100.

Oberauer, K., Lange, E., & Engle, R. W. (2004). Working memory capacity and resistance to interference. *Journal of Memory and Language*, *51*(1), 80–96.

Oberauer, K., Souza, A. S., Druey, M. D., & Gade, M. (2013). Analogous mechanisms of selection and updating in declarative and procedural working memory: experiments and a computational model. *Cognitive Psychology*, *66*(2), 157–211.

Polderman, T. J., de Geus, E. J., Hoekstra, R. A., Bartels, M., van Leeuwen, M., Verhulst, F. C., ... Boomsma, D. I. (2009). Attention problems, inhibitory control, and intelligence index overlapping genetic factors: a study in 9-, 12-, and 18-year-old twins. *Neuropsychology*, *23*(3), 381–391.

Rabbitt, P. (1993). Does it all go together when it goes? The Nineteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology*, *46(A)*(3), 385–434.

Redick, T. S., Calvo, A., Gay, C. E., & Engle, R. W. (2011). Working memory capacity and go/no-go task performance: selective effects of updating, maintenance, and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(2), 308–324.

Roberts, G., & Anderson, M. (2014). Task structure complexity and goal neglect in typically developing children. *Journal of Experimental Child Psychology*, *120*, 59–72.

Roberts, R. J., Jr., & Pennington, B. F. (1996). An interactive framework for examining prefrontal cognitive processes. *Developmental Neuropsychology*, *12*(1), 105–126.

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General*, *123*(4), 374–393.

Salthouse, T. A. (1992). Reasoning and spatial abilities. In F. I. M. Craik, & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition.* New Jersey: Lawrence Erlbaum.

Salthouse, T. A., Atkinson, T. M., & Berish, D. E. (2003). Executive functioning as a potential mediator of age-related cognitive decline in normal adults. *Journal of Experimental Psychology: General*, *132*(4), 566–594.

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: a review of 50 years of research. *Personality and Individual Differences*, *44*(3), 535–551.

Süß, H., Oberauer, K., Wittmann, W. W., Wilhelm, O., & Schulze, R. (2002). Working-memory capacity explains reasoning ability—and a little bit more. *Intelligence*, *30*(3), 261–288.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Unsworth, N., & Engle, R. W. (2007a). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132.

Unsworth, N., & Engle, R. W. (2007b). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychological Bulletin*, *133*(6), 1038–1066.

Unsworth, N., & Spillers, G. J. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language*, *62*(4), 392–406.

Unsworth, N., Redick, T. S., Lakey, C. E., & Young, D. L. (2010). Lapses in sustained attention and their relation to executive control and fluid abilities: an individual differences investigation. *Intelligence*, *38*(1), 111–122.